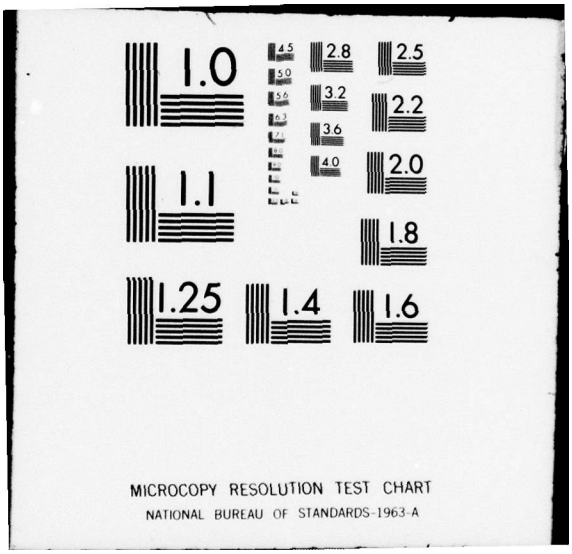
AD-A058 633 GEORGE WASHINGTON UNIV WASHINGTON D C INST FOR MANAG--ETC F/G 12/1
NONLINEAR PROGRAMMING SENSITIVITY ANALYSIS RESULTS USING STRONG--ETC(U)
JUN 78 A V FIACCO DAAG29-76-G-0150
UNCLASSIFIED SERIAL-T-377 ARO-13622.4-M NL

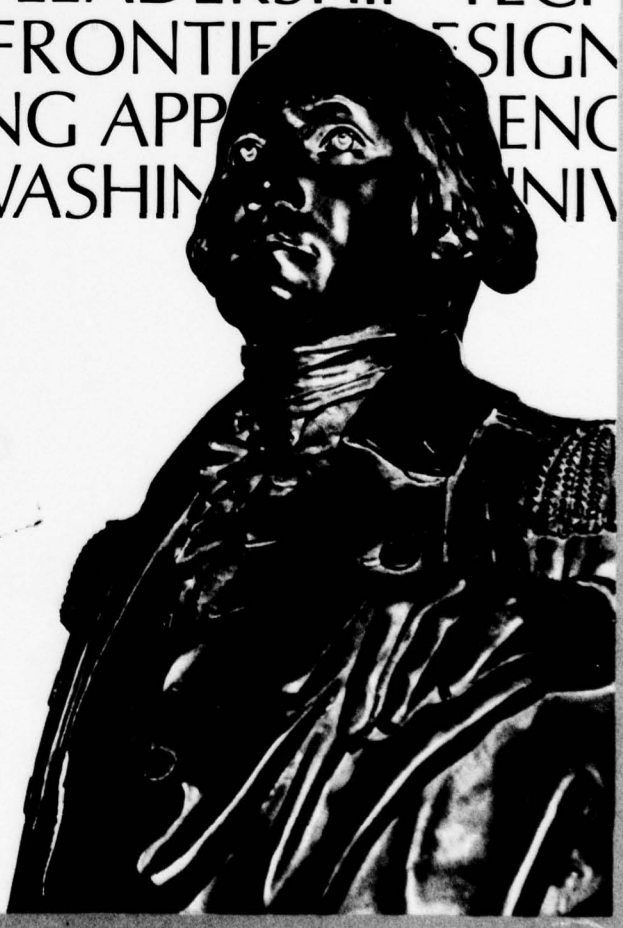
END
DATE FILMED
11-78
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

THE GEORGE WASHINGTON UNIVERSITY

12

STUDENTS FACULTY STUDY RESEARCH DEVELOPMENT FUTURE CAREER CREATIVITY COMMUNITY LEADERSHIP TECHNOLOGY FRONTIE[illegible]SIGN ENGINEERING APP[illegible]ENC GEORGE WASHIN[illegible]NIV



INSTITUTE FOR MANAGEMENT SCIENCE AND ENGINEERING
SCHOOL OF ENGINEERING AND APPLIED SCIENCE

(6) NONLINEAR PROGRAMMING SENSITIVITY ANALYSIS RESULTS USING STRONG SECOND ORDER ASSUMPTIONS.

by

(10) Anthony V. Fiacco

(11) 26 June 78

(9) Technical rept.,

(12) 41 p.

(14) Serial-T-377
26 June 1978

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

(18) ARO
(19) 13622.4-M

(15) U.S. Army Research Office - Durham
Contract Number DAAG29-76-G-0150

406 743

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: ~~T-377~~ T-377 / 2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): NONLINEAR PROGRAMMING SENSITIVITY ANALYSIS RESULTS USING STRONG SECOND ORDER ASSUMPTIONS	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL SCIENTIFIC 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): ANTHONY V. FIACCO	8. CONTRACT OR GRANT NUMBER(s): DAAG-29-76-G-0150
9. PERFORMING ORGANIZATION NAME AND ADDRESS: THE GEORGE WASHINGTON UNIVERSITY INSTITUTE FOR MANAGEMENT SCIENCE & ENGINEERING WASHINGTON, D. C. 20052	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. A. Army Research Office Box 12211 Research Triangle Park, N. S. 27709	12. REPORT DATE: 26 June 1978 13. NUMBER OF PAGES: 36
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): NONE 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF THIS REPORT IS UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

NONLINEAR PROGRAMMING	KUHN-TUCKER TRIPLE
PARAMETRIC PROGRAMMING	OPTIMAL VALUE FUNCTION
SENSITIVITY ANALYSIS	DERIVATIVES
SECOND ORDER ASSUMPTIONS	ALGORITHMIC CALCULATION

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Invoking conditions utilized to obtain numerous "ideal" results in nonlinear programming, this paper summarizes the development of a basis for calculating the first partial derivatives of a Kuhn-Tucker triple and the first and second partial derivatives of the optimal value function, with respect to problem parameters. In the context of prior results, a simpler but much more general derivation of the Kuhn-Tucker triple derivatives is presented, and a more concise formula for the Hessian of the —> next page

(Continued)

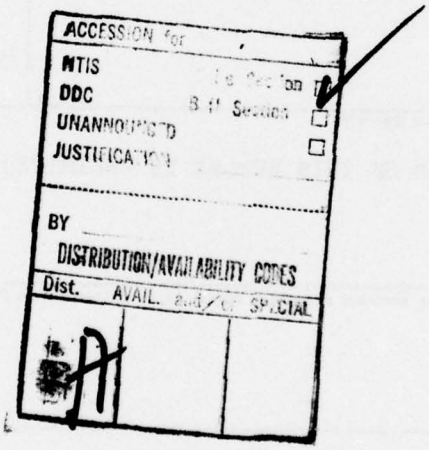
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 |

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract, Continued

optimal value function is given. Particularizations to the problems with right-hand-side constraint perturbations, no constraint perturbations and no constraints follow easily and are briefly treated. Further extensions and applications are indicated.



THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics

Abstract
of
Serial T-377
26 June 1978

NONLINEAR PROGRAMMING SENSITIVITY ANALYSIS RESULTS USING STRONG SECOND ORDER ASSUMPTIONS

by

Anthony V. Fiacco

Invoking conditions utilized to obtain numerous "ideal" results in nonlinear programming, this paper summarizes the development of a basis for calculating the first partial derivatives of a Kuhn-Tucker triple and the first and second partial derivatives of the optimal value function, with respect to problem parameters. In the context of prior results, a simpler but much more general derivation of the Kuhn-Tucker triple derivatives is presented, and a more concise formula for the Hessian of the optimal value function is given. Particularizations to the problems with right hand side constraint perturbations, no constraint perturbations, and no constraints follow easily and are briefly treated. Further extensions and applications are indicated.

TABLE OF CONTENTS

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

NONLINEAR PROGRAMMING SENSITIVITY ANALYSIS RESULTS USING STRONG SECOND ORDER ASSUMPTIONS

by

Anthony V. Fiacco

1. Introduction

Fiacco [15] recently obtained a theoretical basis for locally characterizing the differentiability properties of a local solution and the associated Lagrange multipliers for a large class of nonlinear programming problems with respect to general parametric variations and established the use of a penalty function method to estimate the parameter partial derivatives. Independently, Robinson [22] obtained closely related characterizations of the continuity properties of Kuhn-Tucker points, including bounds on these quantities, and applied his results to derive convergence rates for a family of nonlinear programming algorithms. These sensitivity results are generalizations of a result presented in Fiacco and McCormick [16, Theorem 6] for a particular class of parametric perturbations.

Based on the results of [15], Armacost and Fiacco [4] obtained general expressions for the first and second derivatives of the optimal value function and gave concise expressions for the first derivatives of a Kuhn-Tucker triple, along with approximations of these quantities by way of penalty function calculations. Subsequently, Armacost and Fiacco [5]

particularized these results for right hand side constraint perturbations, and Armacost [2] and Armacost and Fiacco [7] showed the applicability of other (than the usual penalty function) methods in calculating sensitivity information as a byproduct of normal algorithmic calculations. In particular, the approach was shown to apply readily to exact penalty function and augmented Lagrangian algorithms. Buys and Gonin [13] have also independently shown the results for the latter.

The approach given in [15] was implemented computationally by Armacost and Mylander [8]. Computational experience has been reported by Armacost and Fiacco [3], [6], and Armacost [2], [1].

This paper gives a concise summary of the development of the referenced results, drawing heavily on the material presented by Armacost and Fiacco in [4]. However, the derivation of the expression for the derivatives of a Kuhn-Tucker triple is appreciably simplified, while a much more general formula for the derivatives is obtained, and a more concise form is given for the optimal value Hessian. Also, for completeness, we particularize the results to the problem with right hand side constraint perturbations, to the problem with constraints not involving the parameters, and to the unconstrained parametric problem.

In Section 3 we give the relevant basic results. In Section 4 these are applied to develop first and second order changes in the optimal value function of a general class of parametric nonlinear programming problems, with respect to general parametric variations. The usual Lagrange multiplier sensitivity result and the indicated result given in Fiacco and McCormick [16] are obtained as particular instances. In Section 5 we derive a general expression and a number of formulae, depending on problem structure for computing the first partial derivatives of a Kuhn-Tucker triple. Section 6 indicates several extensions and gives a brief discussion of applications.

2. Notation, Preliminaries, and the Problem

The following conventions will be used throughout the paper. E^n denotes the usual n-dimensional Euclidean space. If $x \in E^n$, then x is an $n \times 1$ (column) vector in E^n. The superscript T denotes transposition. The symbols ∇_x and ∇_x^2 denote the gradient and Hessian, respectively, the subscript denoting the variables with respect to which the derivatives are taken. The gradient of a scalar valued function is assumed to be a row vector. Thus, if $f: E^n \times E^h \to E^1$ is once differentiable in E^n, then $\nabla_x f(x,\varepsilon) = [\partial f(x,\varepsilon)/\partial x_1, \ldots, \partial f(x,\varepsilon)/\partial x_n]$, a $1 \times n$ vector. If f is twice differentiable in x, then $\nabla_x^2 f \equiv \nabla_x(\nabla_x f^T)$ denotes the $n \times n$ Hessian matrix of $f(x,\varepsilon)$ with respect to x, whose ij-th element is given by $\partial^2 f(x,\varepsilon)/\partial x_j \partial x_i$ for $i,j = 1,\ldots,n$. Consistent with this, if $g: E^n \times E^h \to E^m$ is a vector function whose components $g_i(x,\varepsilon)$ are differentiable in x, then $\nabla_x g(x,\varepsilon)$ denotes the Jacobian of g with respect to x, an $m \times n$ matrix whose i-th row is given by $\nabla_x g_i(x,\varepsilon)$, $i = 1,\ldots,m$. Differentiation with respect to ε is denoted similarly, of course. Additionally, since there is ample occasion to differentiate such quantities as $f[x(\varepsilon),\varepsilon]$ with respect to ε, the notation $\partial f/\partial \varepsilon \equiv (\partial f/\partial \varepsilon_1, \ldots, \partial f/\partial \varepsilon_h)$ is introduced to indicate and emphasize partial differentiation with respect to the "independent" variable ε only as it appears explicitly. Thus, application of the chain rule for differentiation yields $\nabla_\varepsilon f[x(\varepsilon)\varepsilon] = \nabla_x f \, \nabla_\varepsilon x(\varepsilon) + \partial f/\partial \varepsilon$, where f and $\partial f/\partial \varepsilon$ are evaluated at $x(\varepsilon)$. Analogous to the gradient, if $g[x(\varepsilon),\varepsilon]$ is an m dimensional differentiable vector function on $E^n \times E^h$, then $\partial g/\partial \varepsilon$ is an m by h matrix where i-th row is $\partial g_i/\partial \varepsilon$, evaluated at $x(\varepsilon)$. Arguments of functions are often omitted to simplify the notation, when it is felt that no ambiguity will result.

Turning now to the problem of interest, consider the problem of obtaining a local solution $x(\varepsilon)$ of

$$\begin{aligned} &\text{minimize} \quad f(x,\varepsilon) \\ &x \in E^n \\ &\text{subject to} \quad g_i(x,\varepsilon) \geq 0 , \quad i = 1,\ldots,m , \\ &\qquad\qquad h_j(x,\varepsilon) = 0 , \quad i = 1,\ldots,p , \end{aligned} \qquad P(\varepsilon)$$

where ε is a parameter in E^h and f , g_i , h_j are real valued functions.

We are interested in studying the behavior of a local solution $x(\varepsilon)$ and its associated (optimal) Lagrange multipliers $u(\varepsilon)$, $w(\varepsilon)$ for small changes in the parameter vector ε , near a specified value of the parameter. We are also interested in the "optimal value" $f[x(\varepsilon),\varepsilon]$ of $P(\varepsilon)$. Without loss of generality, we assume the specified value is $\varepsilon = 0$.

In this paper, we assume conditions strong enough to guarantee the existence and differentiability of $x(\varepsilon)$, $u(\varepsilon)$, $w(\varepsilon)$ near $\varepsilon = 0$. Key conditions are the well known second order sufficient conditions for a locally unique solution of Problem P(0). These may be found in Fiacco and McCormick [16,Theorem 4] and in numerous current books and papers. For, completeness, we state them here, in the context of Problem P(0).

Define the Lagrangian of $P(\varepsilon)$ as

$$L(x,u,w,\varepsilon) \equiv f(x,\varepsilon) - \sum_{i=1}^{m} u_i g_i(x,\varepsilon) + \sum_{j=1}^{p} w_j h_j(x,\varepsilon) \qquad (2.1)$$

Second order conditions are intended to mean conditions based on the assumption that the problem functions are twice continuously differentiable. The second order sufficient conditions are said to hold for Problem P(0) at a point x^* if

(i) x^* is a feasible point of P(0) ,

(ii) there exist $u^* = (u_1^*,\ldots,u_m^*)^T$ and $w^* = (w_1^*,\ldots,w_p^*)^T$ such that $\nabla_x L(x^*,u^*,w^*,0) = 0$, $u_i^* g_i(x^*) = 0$ and $u_i^* \geq 0$ for $i = 1,\ldots,m$, and

(iii) $Z^T \nabla_x^2 L(x^*,u^*,w^*,0)Z > 0$ for every nonzero vector $Z \varepsilon E^n$ satisfying $Z^T \nabla_x g_i(x^*) \geq 0$ for all i such that $g_i(x^*) = 0$, $Z^T \nabla g_i(x^*) = 0$ for all i such that $g_i(x^*) = 0$ and $u_i^* > 0$, and $Z^T \nabla h_j(x^*) = 0$ for all $j = 1,\ldots,p$.

If these conditions hold, then it follows that x^* is an isolated (i.e., locally unique) local minimum of P(0), with associated (not necessarily unique) "optimal" Lagrange multipliers u^*,w^*. Conditions (i) and (ii) require first derivatives only and are known as the (first order) Kuhn-Tucker conditions. If (x^*,u^*,w^*) satisfies (i) - (iii), it will be called a "Kuhn-Tucker triple."

A few facts concerning these conditions should be noted, since they have reference to interpreting the results that follow. If no constraints are present, then if we suppress reference to the terms associated with the constraints, the conditions reduce to

$$\text{(i)} \quad x^* \varepsilon E^n ,$$

$$\text{(ii)} \quad \nabla_x f(x^*) = 0 \quad \text{and}$$

$$\text{(iii)} \quad Z^T \nabla^2 f(x^*) Z > 0$$

for every nonzero vector $Z \varepsilon E^n$, and the conclusion becomes x^* is an isolated local minimum of $f(x)$. Thus, the conditions reduce to the well known second order sufficient conditions for an unconstrained local minimum of $f(x)$. If the Problem P(0) is linear, then $\nabla_x^2 L \equiv 0$. In this case, the sufficient conditions as stated cannot possibly hold, unless there are no nonzero vectors Z satisfying the stipulated requirements, in which instance there is no inconsistency in the given conditions. It can be shown that the conclusion that x^* is an isolated local minimum still follows under these circumstances. It may be observed that the nonexistence of any vector Z as stated implies, in particular, that there is no nonzero vector orthogonal to all the binding constraint gradients, hence, there must be

n binding constraints whose gradients are linearly independent. This implies that x^* is a vertex of the polyhedron defining the constraint set, consistent with the conclusion that x^* is an isolated local minimum. Whether the problem is linear or not, if there is no nonzero vector Z satisfying the given conditions, then the second order requirement (iii) is satisfied in a logical sense, and in fact it can be shown that the first order conditions (i) and (ii) are then sufficient to conclude that x^* is an isolated local minimum. Finally, if $P(0)$ is a convex programming problem [16, Chapter 6] then, since a local solution of a convex problem is global, the sufficient conditions imply that x^* is the unique global solution.

3. First Order Changes in a Kuhn-Tucker Triple

The following result provides the basis for the development that will be given here.

THEOREM 3.1 (Characterization of a Differentiable Kuhn-Tucker Triple, Fiacco [15, Theorem 2.1])

If

(i) the functions defining $P(\varepsilon)$ are twice continuously differentiable in (x,ε) in a neighborhood of $(x^*,0)$;

(ii) the second order sufficient conditions (Fiacco and McCormick [16, Theorem 4]) for a local minimum of $P(0)$ hold at x^* with associated Lagrange multipliers u^* and w^* ;

(iii) the gradients $\nabla_x g_i(x^*,0)$ (for i such that $g_i(x^*,0) = 0$) and $\nabla_x h_j(x^*,0)$ (all j) are linearly independent; and

(iv) strict complimentarity holds at $(x^*,0)$ with respect to u^*, i.e., $u_i^* > 0$ when $g_i(x^*,0) = 0$ $(i=1,\ldots,m)$;

then

(a) x^* is a local isolated minimizing point of $P(0)$ and the associated Lagrange multipliers u^* and w^* are unique;

(b) for ε in a neighborhood of 0, there exists a unique once continuously differentiable vector function $y(\varepsilon) = [x(\varepsilon)^T, u(\varepsilon)^T, w(\varepsilon)^T]^T$ (where T denotes transposition) satisfying the second order sufficient conditions for a local minimum of Problem $P(\varepsilon)$ such that $y(0) = (x^{*T}, u^{*T}, w^{*T})^T$ and, hence, $x(\varepsilon)$ is a locally unique minimum of $P(\varepsilon)$ with associated Lagrange multipliers $u(\varepsilon)$ and $w(\varepsilon)$; and

(c) strict complementarity and linear independence of the binding constraint gradients hold at $x(\varepsilon)$ for ε near 0.

The conditions of the theorem will be assumed throughout the remainder of the paper.

When $y(\varepsilon)$ is available, $\nabla_\varepsilon y(\varepsilon) = (\nabla_\varepsilon x(\varepsilon)^T, \nabla_\varepsilon u(\varepsilon)^T, \nabla_\varepsilon w(\varepsilon)^T)^T$ (an $(n+m+p)$ by k matrix) can be calculated by noting that the theorem implies the satisfaction of the Kuhn-Tucker conditions for $P(\varepsilon)$ at $y(\varepsilon)$ near $\varepsilon = 0$, i.e.,

$$\begin{aligned} \nabla_x L[x(\varepsilon), u(\varepsilon), w(\varepsilon), \varepsilon] &= 0 , \\ u_i(\varepsilon) g_i[x(\varepsilon), \varepsilon] &= 0 , \quad i = 1,\ldots,m , \\ h_j[x(\varepsilon), \varepsilon] &= 0 , \quad i = 1,\ldots,p . \end{aligned} \tag{3.1}$$

Since near $\varepsilon = 0$ the Jacobian, $M(\varepsilon)$, of this system with respect to (x,u,w) is nonsingular under the given assumptions, the total derivative of the system with respect to ε is well defined and must equal zero. This yields

$$M(\varepsilon)\nabla_\varepsilon y(\varepsilon) = N(\varepsilon) \tag{3.2}$$

where $N(\varepsilon)$ is the negative of the Jacobian of the Kuhn-Tucker system with respect to ε, and hence

$$\nabla_\varepsilon y(\varepsilon) = M(\varepsilon)^{-1} N(\varepsilon) \ . \tag{3.3}$$

If there are no constraints present in $P(0)$ then Theorem 3.1 reduces to the statement that if $f(x,\varepsilon)$ is twice continuously differentiable in (x,ε) near $(x^*,0)$, and if $\nabla_x f(x^*,0) = 0$ and $\nabla_x^2 f(x^*,0)$ is positive definite, then x^* is a local isolated unconstrained minimum of $f(x^*,0)$, and there exists near $\varepsilon = 0$ a unique once continuously differentiable function $x(\varepsilon)$ satisfying $\nabla_x f[x(\varepsilon),\varepsilon] = 0$, with $\nabla_x^2 f[x(\varepsilon),\varepsilon]$ positive definite and such that $x(0) = x^*$. Equation (3.2) becomes

$$\nabla_x^2 f[x(\varepsilon),\varepsilon] \ \nabla_\varepsilon x(\varepsilon) + \frac{\partial}{\partial \varepsilon} \nabla_x f^T[x(\varepsilon),\varepsilon] = 0 \tag{3.4}$$

and hence (3.3) becomes

$$\nabla_\varepsilon x(\varepsilon) = \nabla_x^2 f[x(\varepsilon),\varepsilon]^{-1} \frac{\partial}{\partial \varepsilon} \nabla_x f^T[x(\varepsilon),\varepsilon] \ . \tag{3.5}$$

These calculations will be pursued in some detail in Section 5. First, however, we give several results characterizing the optimal value function $f[x(\varepsilon),\varepsilon]$ that follow immediately from Theorem 3.1.

4. First and Second Order Changes in the Optimal Value Function

Because of important connections with Lagrange multipliers and duality theory, first order changes in the optimal value function have traditionally been analyzed with respect to variations in the "right hand side" of the constraints. An extension to perturbations (of <u>all</u> problem functions) that are <u>linear</u> in the problem parameters was obtained by Fiacco and McCormick [16]. Buys [12] also derives <u>second order</u> changes, in connection with an analysis of the behavior of the optimal value of an associated augmented Lagrangian function. Here, under the assumptions of Theorem 3.1, it is shown how first and second order results follow immediately for the general class of parametric variations being considered. The referenced results are obtained as special cases.

Let $y(\varepsilon) = [x(\varepsilon)^T, u(\varepsilon)^T, w(\varepsilon)^T]^T$ be a Kuhn-Tucker triple, where $x(\varepsilon)$ solves Problem $P(\varepsilon)$ for ε near 0. The "optimal value function" is defined as

$$f^*(\varepsilon) \equiv f[x(\varepsilon),\varepsilon] , \tag{4.1}$$

and the "optimal value Lagrangian" as

$$L^*(\varepsilon) = L[x(\varepsilon),u(\varepsilon),w(\varepsilon),\varepsilon] . \tag{4.2}$$

THEOREM 4.1 (First and Second Order Changes in the Optimal Value Function of Problem $P(\varepsilon)$, Armacost and Fiacco [4, Theorem 3])

If the conditions of Theorem 3.1 hold for Problem $P(\varepsilon)$, then, in a neighborhood of $\varepsilon = 0$,

$$\text{(a)} \quad f^*(\varepsilon) = L^*(\varepsilon) , \tag{4.3}$$

$$\text{(b)} \quad \nabla_\varepsilon f^*(\varepsilon) = \partial L/\partial\varepsilon = \partial f/\partial\varepsilon - \sum_{i=1}^{m} u_i(\varepsilon)[\partial g_i/\partial\varepsilon] + \sum_{j=1}^{p} w_j(\varepsilon)[\partial h_j/\partial\varepsilon]$$

$$= \partial f/\partial\varepsilon - u(\varepsilon)^T(\partial g/\partial\varepsilon) + w(\varepsilon)^T(\partial h/\partial\varepsilon) , \tag{4.4}$$

and hence also

$$\text{(c)} \quad \nabla_\varepsilon^2 f^*(\varepsilon) = \nabla_\varepsilon[(\partial L/\partial\varepsilon)^T]$$

$$= \nabla_x[(\partial L/\partial\varepsilon)^T]\, \nabla_\varepsilon x(\varepsilon) - \sum_{i=1}^{m} [\partial g_i/\partial\varepsilon]^T \nabla_\varepsilon u_i(\varepsilon)$$

$$+ \sum_{j=1}^{p} [\partial h_j/\partial\varepsilon]^T \nabla_\varepsilon w_j(\varepsilon) + \frac{\partial^2 L}{\partial\varepsilon^2} \tag{4.5}$$

Proof: Recall that in a neighborhood of $\varepsilon = 0$, $u_i(\varepsilon)g_i[x(\varepsilon),\varepsilon] \equiv 0$, $i = 1,\ldots,m$, strict complementary slackness holds, $h_j[x(\varepsilon),\varepsilon] \equiv 0$, $j = 1,\ldots,p$, and $y(\varepsilon) = [x(\varepsilon)^T, u(\varepsilon)^T, w(\varepsilon)^T]^T \in C^1$. It follows immediately that

$$f[x(\varepsilon),\varepsilon] \equiv L[x(\varepsilon),u(\varepsilon),w(\varepsilon),\varepsilon] , \tag{4.6}$$

yielding (a). Furthermore, we can differentiate (4.6) to obtain

$$\nabla_\varepsilon f^*(\varepsilon) = \nabla_\varepsilon f[x(\varepsilon),\varepsilon] = \nabla_\varepsilon L^*(\varepsilon) = \nabla_\varepsilon L[x(\varepsilon),u(\varepsilon),w(\varepsilon),\varepsilon]$$

$$= \nabla_x L \nabla_\varepsilon x(\varepsilon) + \nabla_u L \nabla_\varepsilon u(\varepsilon) + \nabla_w L \nabla_\varepsilon w(\varepsilon) + \partial L/\partial\varepsilon \ , \tag{4.7}$$

where L is evaluated as in (4.6).

Since the Kuhn-Tucker conditions hold at $y(\varepsilon)$, it follows that $\nabla_x L = 0$. Complementary slackness implies $u_i(0)$ or $g_i[x(0),0] = 0$, $i = 1,\ldots,m$. Strict complementary slackness, continuity and differentiability then imply one of two consequences, respectively:

(i) $g_i[x(0),0] > 0$, implying $g_i[x(\varepsilon),\varepsilon] > 0$ for ε near 0, implying $u_i(\varepsilon) \equiv 0$, implying $\nabla_\varepsilon u_i(\varepsilon) \equiv 0$; or

(ii) $u_i(0) > 0$, implying $u_i(\varepsilon) > 0$ for ε near 0, implying $g_i[x(\varepsilon),\varepsilon] \equiv 0$.

From this, it follows that $\nabla_u L \nabla_\varepsilon u(\varepsilon) = (-g_1[x(\varepsilon),\varepsilon],\ldots,-g_m[x(\varepsilon),\varepsilon])$ $\nabla_\varepsilon u(\varepsilon) = 0$. Also, since $h_j[x(\varepsilon),\varepsilon] \equiv 0$ for ε near 0, $\nabla_w L \nabla_\varepsilon w(\varepsilon) = (h_1[x(\varepsilon),\varepsilon],\ldots,h_p[x(\varepsilon),\varepsilon])\nabla_\varepsilon w(\varepsilon) = 0$. We therefore conclude from (4.7) that $\nabla_\varepsilon f^*(\varepsilon) = \partial L/\partial\varepsilon$ for ε near 0, proving (b).

Differentiation with respect to ε of the result obtained in (b) gives

$$\nabla_\varepsilon^2 f^*(\varepsilon) = \nabla_x[\partial L/\partial\varepsilon^T]\nabla_\varepsilon x(\varepsilon) + \nabla_u[\partial L/\partial\varepsilon^T]\nabla_\varepsilon u(\varepsilon)$$

$$+ \nabla_w[\partial L/\partial\varepsilon^T]\nabla_\varepsilon w(\varepsilon) + \partial^2 L/\partial\varepsilon^2 \ .$$

Calculation of the derivatives yields (c).

To be perfectly clear about what is involved in calculating this Hessian, we write Equation (4.5) in terms of the original problem functions. We have

$$\nabla_\varepsilon^2 f^*(\varepsilon) = \nabla_x \left\{ \left(\frac{\partial f}{\partial \varepsilon} \right)^T - \sum_{i=1}^{m} u_i(\varepsilon) \left(\frac{\partial g_i}{\partial \varepsilon} \right)^T + \sum_{j=1}^{p} w_j(\varepsilon) \left(\frac{\partial h_j}{\partial \varepsilon} \right)^T \right\} \cdot \nabla_\varepsilon x(\varepsilon)$$

$$- \sum_{i=1}^{m} \left\{ \left(\frac{\partial g_i}{\partial \varepsilon} \right)^T \cdot \nabla_\varepsilon u_i(\varepsilon) + \sum_{j=1}^{p} \left(\frac{\partial h_j}{\partial \varepsilon} \right)^T \cdot \nabla_\varepsilon w_j(\varepsilon) \right\} . \qquad (4.8)$$

$$+ \left\{ \frac{\partial^2 f}{\partial \varepsilon^2} - \sum_{i=1}^{m} u_i(\varepsilon) \frac{\partial^2 g_i}{\partial \varepsilon^2} + \sum_{j=1}^{p} w_j(\varepsilon) \frac{\partial^2 h_j}{\partial \varepsilon^2} \right\}$$

Equation (4.4) reduces to previously established results when certain problem structures are considered. The first corollary gives the well-known "Lagrange multiplier sensitivity result," and also establishes relations for the Hessian of the optimal value function taken with respect to the right hand side of the constraints. For this case, Problem $P(\varepsilon)$ reduces to

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \geq \varepsilon_i \quad , \quad i = 1,\ldots,m \, , \qquad R(\varepsilon) \\ &&& h_j(x) = \varepsilon_{j+m} \, , \quad j = 1,\ldots,p \, . \end{aligned}$$

COROLLARY 4.1 (First and Second Order Changes in the Optimal Value Function for Right Hand Side Perturbations, Armacost and Fiacco [4, Corollary 3.1])

If

(i) the functions defining $R(\varepsilon)$ in a neighborhood of $\varepsilon = 0$ are twice continuously differentiable in x, in a neighborhood of x^*, and

(ii) conditions (ii) - (iv) of Theorem 3.1 hold,

then, in a neighborhood of $\varepsilon = 0$,

$$\text{(a)} \quad \nabla_\varepsilon f^*(\varepsilon)^T = \begin{bmatrix} u(\varepsilon) \\ -w(\varepsilon) \end{bmatrix} , \text{ and}$$

$$\text{(b)} \quad \nabla_\varepsilon^2 f^*(\varepsilon) = \begin{bmatrix} \nabla_\varepsilon u(\varepsilon) \\ -\nabla_\varepsilon w(\varepsilon) \end{bmatrix} .$$

Proof: We let $f(x,\varepsilon) = f(x)$; $g_i(x,\varepsilon) = g_i(x) - \varepsilon_i$, $i = 1,\ldots,m$, $h_j(x,\varepsilon) = h_j(x) - \varepsilon_{j+m}$, $j = 1,\ldots,p$, and apply the results of Theorem 4.1.

A second corollary is a generalization of the result established by Fiacco and McCormick [16, Theorem 6] for the problem

$$\begin{aligned} \text{minimize} \quad & f(x) + \varepsilon_0 a_0(x) \\ \text{subject to} \quad & g_i(x) + \varepsilon_i b_i(x) \geq 0 \quad , \quad i = 1,\ldots,m \, , \qquad \tilde{R}(\varepsilon) \\ & h_j(x) + \varepsilon_{j+m} c_j(x) = 0 \, , \quad j = 1,\ldots,p \, . \end{aligned}$$

COROLLARY 4.2 (First Order Changes, Fiacco and McCormick [16, Theorem 6], Armacost and Fiacco [4, Corollary 3.2], and Second Order Changes in the Optimal Value Function for Perturbations Linear in the Parameters)

If

(i) the functions of Problem $\tilde{R}(\varepsilon)$ in a neighborhood of $\varepsilon = 0$ are twice continuously differentiable in x , in a neighborhood of x^* , and

(ii) conditions (ii) - (iv) of Theorem 3.1 hold,

then, in a neighborhood of $\varepsilon = 0$,

$$\text{(a)} \quad \nabla_\varepsilon f^*(\varepsilon)^T = \begin{bmatrix} a_0[x(\varepsilon)] \\ -u_1(\varepsilon) b_1[x(\varepsilon)] \\ \vdots \\ -u_m(\varepsilon) b_m[x(\varepsilon)] \\ w_1(\varepsilon) c_1[x(\varepsilon)] \\ \vdots \\ w_p(\varepsilon) c_p[x(\varepsilon)] \end{bmatrix} = \begin{bmatrix} a_0[x(\varepsilon)] \\ -B[x(\varepsilon)]u(\varepsilon) \\ C[x(\varepsilon)]w(\varepsilon) \end{bmatrix} \, , \text{ and}$$

$$\text{(b)}\quad \nabla_\varepsilon^2 f^*(\varepsilon) = \begin{bmatrix} \nabla_x a_0[x(\varepsilon)]\ \nabla_\varepsilon x(\varepsilon) \\ -U(\varepsilon)\ \nabla_x b[x(\varepsilon)]\ \nabla_\varepsilon x(\varepsilon) - B[x(\varepsilon)]\ \nabla_\varepsilon u(\varepsilon) \\ W(\varepsilon)\ \nabla_x c[x(\varepsilon)]\ \nabla_\varepsilon x(\varepsilon) + C[x(\varepsilon)]\ \nabla_\varepsilon w(\varepsilon) \end{bmatrix} ,$$

where

$$U \equiv \operatorname{diag}(u_i)\ ,\ B \equiv \operatorname{diag}(b_i)\ ,\ i = 1,\ldots,m\ ,$$

$$W \equiv \operatorname{diag}(w_j)\ ,\ C \equiv \operatorname{diag}(c_j)\ ,\ j = 1,\ldots,p\ ,$$

$b = (b_1,\ldots,b_m)^T$ and $c = (c_1,\ldots,c_p)^T$.

Proof: We let $f(x,\varepsilon) = f(x) + \varepsilon_0 a_0(x)$, $g_i(x,\varepsilon) = g_i(x) + \varepsilon_i b_i(x)$, $i = 1,\ldots,m$, $h_j(x,\varepsilon) = h_j(x) + \varepsilon_{j+m} c_j(x)$, $j = 1,\ldots,p$, and apply the results of Theorem 4.1, having verified as in Corollary 4.1 that the conditions of Theorem 3.1 are satisfied. In particular, with * denoting evaluation at $\varepsilon = 0$, we have that

$$\nabla_\varepsilon f^*(0) = (a_0^*, -u_1^* b_1^*, \ldots, -u_m^* b_m^*, w_1^* c_1^*, \ldots, w_p^* c_p^*)\ ,$$

the result obtained in [16, Theorem 6]. Conclusion (b) follows from differentiation of (a).

A third corollary summarizes the well known results that follow in the absence of constraints. Note that under the given conditions, the corollary also applies if constraints are present in P(0) , but are not binding at x^* .

COROLLARY 4.3 (First and Second Order Changes in the Unconstrained Optimal Value Function)

If $f(x,\varepsilon)$ is twice continuously differentiable in (x,ε) near $(x^*,0)$ and if $\nabla_x f(x^*,0) = 0$ and $\nabla_x^2 f(x^*,0)$ is positive definite, then, in a neighborhood of $\varepsilon = 0$,

$$\text{(a)} \quad \nabla_\varepsilon f^*(\varepsilon) = \frac{\partial f}{\partial \varepsilon} \ ,$$

and

$$\text{(b)} \quad \nabla_\varepsilon^2 f^*(\varepsilon) = \frac{\partial}{\partial \varepsilon} (\nabla_x f^T) \nabla_\varepsilon x(\varepsilon) + \frac{\partial^2 f}{\partial \varepsilon^2} \ ,$$

or equivalently,

$$\text{(b)}' \quad \nabla_\varepsilon^2 f^*(\varepsilon) = - \left[\frac{\partial}{\partial \varepsilon} (\nabla_x f^T)\right]^T \nabla^2 f^{-1} \left[\frac{\partial}{\partial \varepsilon} (\nabla_x f^T)\right] + \frac{\partial^2 f}{\partial \varepsilon^2} \ .$$

Proof: Suppress reference to all terms involving the constraints in Theorem 3.1 to conclude the existence of $x(\varepsilon)$ satisfying the conditions stated in the conclusions of the theorem and further interpreted in the paragraph just preceding Equation (3.4). Differentiate $f^*(\varepsilon) \equiv f[x(\varepsilon),\varepsilon]$ with respect to ε and use $\nabla_x f[x(\varepsilon),\varepsilon] = 0$ to obtain (a). Differentiate $\nabla_\varepsilon f[x(\varepsilon),\varepsilon] = 0$ to obtain (b) and use the fact that $\nabla_x^2 f[x(\varepsilon),\varepsilon]$ is positive definite and (3.5) to obtain (b)'.

It is interesting to note that if the constraints of $P(\varepsilon)$ are independent of ε, then application of the results of Theorem 4.1 yields the same expressions for $\nabla_\varepsilon f^*(\varepsilon)$ and $\nabla_\varepsilon^2 f^*(\varepsilon)$ as those obtained above in (a), (b) and (b'), respectively, for the unconstrained problem.

Note from Theorem 4.1 that the values of the optimal value function and its gradient can be calculated once the Kuhn-Tucker triple $y(\varepsilon)$ has been determined. However, in general, the value of the Hessian matrix of the optimal value function requires the determination of both the triple and its first derivatives.

We next examine various aspects of calculating these derivatives. Aside from their use in calculating the optimal value function Hessian, they are of considerable importance in other applications, e.g., in characterizing the stability of the solution subject to perturbation and in providing a first order estimate of Kuhn-Tucker triples of problems involving different values of the parameters, once one such triple has been determined. The analysis leads to a study of the Jacobian $M(\varepsilon)$ of the Kuhn-Tucker system (3.1).

5. Computation of First Order Changes in the Kuhn-Tucker Triple

Our task is to calculate $\nabla_\varepsilon y(\varepsilon)$ for Problem $P(\varepsilon)$ when the conditions of Theorem 3.1 hold. As noted, once $y(\varepsilon)$ is available, $\nabla_\varepsilon y(\varepsilon)$ can be calculated by using (3.2) or (3.3). Appreciable efficiencies in computation can be introduced by analyzing the various possibilities.

Conclusion (c) of Theorem 3.1 implies that, near $\varepsilon = 0$, $u_i(\varepsilon) > 0$ if $u_i(0) > 0$ and, using Assumption (i), we can also conclude that $g_i[x(\varepsilon),\varepsilon] > 0$ for all i such that $g_i[x(0),0] > 0$, which in turn implies that $u_i(\varepsilon) \equiv 0$ whenever $g_i[x(0),0] > 0$. Using these facts essentially allows us to eliminate those terms associated with constaints that are not binding at x^*, and also allows us to divide out the positive $u_i(\varepsilon)$ from the corresponding complimentary slackness equations.

This leads to a considerable simplification of the Kuhn-Tucker conditions (3.1), which must hold near $\varepsilon = 0$. Without loss of generality, assume that the first r inequality constraints are binding. We are thus lead to studying the system,

$$\begin{aligned} \nabla_x L[x(\varepsilon),u(\varepsilon),w(\varepsilon),\varepsilon] &= 0 \\ -\bar{g}[x(\varepsilon),\varepsilon] &= 0 \\ h[x(\varepsilon),\varepsilon] &= 0 , \end{aligned} \tag{5.1}$$

where $\bar{g} = (g_1,\ldots,g_r)^T$ and $h = (h_1,\ldots,h_p)^T$. (The minus sign before $\bar{g}$ leads to notational simplifications.)

It is assumed in the following development that the analysis is confined to a neighborhood of $\varepsilon = 0$ where the conclusions of Theorem 3.1 are valid.

Differentiating (5.1) with respect to ε according to the chain rule yields

$$\bar{M}(\varepsilon)\ \nabla_\varepsilon \bar{y}(\varepsilon) = \bar{N}(\varepsilon) \tag{5.2}$$

where the Jacobians $\bar{M}$ and $-\bar{N}$ of (5.2) with respect to (x,u,w) and ε are, respectively,

$$\bar{M} \equiv \begin{bmatrix} \nabla^2 L & P^T \\ P & 0 \end{bmatrix} \tag{5.3}$$

and

$$-\bar{N} \equiv [(\partial(\nabla_x L^T)/\partial\varepsilon)^T, -(\partial\bar{g}/\partial\varepsilon)^T, (\partial h/\partial\varepsilon)^T]^T , \tag{5.4}$$

where

$$\begin{aligned} P &= (-\nabla_x \bar{g}^T, \nabla_x h^T)^T , \\ \bar{y}(\varepsilon) &= (x(\varepsilon)^T, \bar{u}(\varepsilon)^T, w(\varepsilon)^T)^T , \end{aligned} \tag{5.5}$$

and

$$\bar{u}(\varepsilon) = (u,(\varepsilon),\ldots,u_r(\varepsilon))^T .$$

Under the given conditions, it is known that the Jacobian $M(\varepsilon)$ of (3.1) with respect to (u,v,w) is nonsingular, and hence it follows that $\bar{M}$ defined in (5.3) is nonsingular, for ε near 0 . Thus,

$$\nabla_\varepsilon \bar{y}(\varepsilon) = \bar{M}(\varepsilon)^{-1} \bar{N}(\varepsilon) . \tag{5.6}$$

Clearly, any method for solving the linear system of Equation (5.2) is applicable for calculating $\nabla_\varepsilon \bar{y}(\varepsilon)$, and $\bar{M}$ need not be inverted as in (5.6). However, under the given assumptions, the work involved in calculating $M(\varepsilon)^{-1}$ can be significantly reduced, as will presently become evident.

Consider the various possibilities: (1) $\nabla^2 L^{-1}$ exists; (2) $\nabla_x^2 L = 0$; (3) $r + p = n$ (i.e., there are n binding constraints), or (4) $r + p < n$. Brief reflection will indicate that <u>all</u> possibilities are covered by these conditions (in fact, under the given assumptions, either (3) or (4) must be true, but (1) and (2) are also specified since they introduce further simplifications).

Letting

$$\bar{M}(\varepsilon)^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} , \tag{5.7}$$

where A_{11} is n x n , we can obtain the block components A_{ij} for each case delineated, requiring at most the inversion of an n by n matrix. The first three cases follow readily from a straightforward manipulation of (5.3).

Case 1. $\nabla_x^2 L^{-1}$ exists.

Since the assumptions of Theorem 3.1 guarantee that $[P\nabla_x^2 L^{-1} P^T]^{-1}$ exists, it is easily shown that

$$\begin{aligned} A_{11} &= \nabla_x^2 L^{-1}[I - P^T(P\nabla_x^2 L^{-1}P^T)^{-1}P\nabla_x^2 L^{-1}] \\ A_{12} &= A_{21}^T = \nabla_x^2 L^{-1} P^T [P\nabla_x^2 L^{-1} P^T]^{-1} \\ A_{22} &= -[P\nabla_x^2 L^{-1} P^T]^{-1} . \end{aligned} \tag{5.8}$$

Case 2. $\nabla_x^2 L = 0$.

There are two possible situations: there are $r + p < n$, or $r + p = n$ linearly independent binding constraint gradients. If there are less than n , Assumption (ii) of Theorem 1 is violated (and it is easily seen that $\bar{M}(\varepsilon)^{-1}$ does not exist), so this is not allowed under the present assumptions. When there are n linearly independent binding constraint gradients, then we have a special instance of the third circumstance which follows.

Case 3. There are n linearly independent binding constraint gradients.

The n x n Jacobian P of the n constraints with respect to x must be nonsingular. Hence,

$$A_{11} = 0$$

$$A_{12} = A_{21}^T = P^{-1} \tag{5.9}$$

$$A_{22} = -P^{-T} \nabla_x^2 L \, P^{-1} .$$

Note that if $\nabla_x^2 L^{-1}$ exists, (5.9) follows from (5.8); however, here the existence of $\nabla_x^2 L^{-1}$ is not assumed. Also, the remaining possibility mentioned in Case 2 above gives $A_{11} = A_{22} = 0$ and $A_{12} = A_{21}^T = P^{-1}$. (It may be of interest to note that this last situation characterizes conditions that hold at a nondegenerate solution of a linear programming problem, with n linearly independent binding constraint gradients and $\nabla_x^2 L = 0$.)

Case 4. $r + p < n$ and $\nabla_x^2 L \neq 0$.

This is the least structured, and hence most general, situation that can be encountered under the given assumptions. Many representations of $\bar{M}^{-1}$ are possible, depending on how the data are organized. However, a general representation that is tailored to the assumptions we are making here was obtained by McCormick [20] and will serve our purpose extremely aptly. We follow his development here rather closely.

First, note that assumptions (ii) and (iii) imply that, at (x^*,u^*,w^*) with $\varepsilon = 0$, we must have $Z^T \nabla_x^2 LZ > 0$ for all $Z \neq 0$ such that $PZ = 0$, where $P = (-\nabla_x \bar{g}^T, \nabla_x h^T)^T$, the $(r+p)$ by n matrix defined in (5.5). Hence, if S is any n by $[n-(r+p)]$ matrix that generates the null space of P, then $Z = Sy$ for some y in $E^{n-(r+p)}$ implies that $PZ = 0$ (since $PS=0$) and also that $Z \neq 0$ if $y = 0$ (since we are tacitly assuming that S has full column rank $[n-(r+p)]$) and hence, $Z^T \nabla_x^2 LZ = y^T S^T \nabla^2 LSy > 0$, providing that $y \neq 0$. We conclude that $D = S^T \nabla_x^2 LS$ is a positive definite $[n-(r+p)]$ by $[n-(r+p)]$ matrix. Further, since P has rank $r + p$ by Assumption (iii), an n by $(r+p)$ pseudo-inverse of P

exists, i.e., a rank (r+p) matrix $P^{\#}$ satisfying $PP^{\#}P = P$. These constructs and observations lead to a general representation of the block components A_{ij} of $\bar{M}^{1}$.

As indicated in [20], $P^{\#}$ and S are assumed in practice to be generated by some matrix technique which relates these quantities by the expression $I-P^{\#}P = SW$, where W is some $[n-(r+p)]$ by n matrix. Also, since $PP^{\#} = I$ and (3.1) gives $\nabla_x^T f(x^*,0) = -P^T[(u^*)^T,(w^*)^T]^T$ at optimality, it may be of interest to note that

$$\begin{pmatrix} u^* \\ w^* \end{pmatrix} = -(P^{\#})^T \nabla_x f(x^*,0) .$$

This motivates the widely used estimation $-(P^{\#})^T \nabla_x f$ for the Lagrange multipliers in algorithms involving these constructs.

The result (equivalent to that given in [20]), in forms of the block components A_{ij} of $\bar{M}^{1}$, follows readily and is given by

$$\begin{aligned} A_{11} &= SD^{-1}S^T , \\ A_{12} &= A_{21}^T = [I-A_{11}\nabla_x^2 L]P^{\#} , \\ A_{22} &= -A_{21}\nabla_x^2 LP^{\#} . \end{aligned} \tag{5.10}$$

There are many good techniques currently available for calculating S and $P^{\#}$, motivated by various numerical efficiency, stability and algorithmic considerations [20]. We mention two here for completeness and because they are precisely tailored to the calculations associated with two important families of mathematical programming algorithms, reduced gradient and projected gradient type algorithms.

The first technique is associated with the reduced gradient or variable reduction type algorithms for nonlinear programming, and is a crucial part of the simplex method for linear programming. It is based on the simple observation that the linear independence assumption implies the

existence of an (r+p) by (r+p) nonsingular submatrix P_D of P. Assuming for simplicity that the first $r + p$ columns of P are linearly independent, we can partition P as $P = (P_D, P_I)$, where P_I is an (r+p) by [n-(r+p)] matrix. This induces a natural decomposition of the variable $x = (x_D^T, x_I^T)^T$, and since $\bar{g}(x^*,0) = 0$ and $h(x^*,0) = 0$, allows application of the implicit function theorem to conclude that there exists a twice differentiable vector function $x_D(x_I)$ such that $\bar{g}[x_D(x_I),x_I,0] \equiv 0$ and $h[x_D(x_I),x_I,0] \equiv 0$ near $x_I = x_I^*$, and $(x_D(x_I^*),x_I^*) = x^*$. The x_D and x_I may be thought of as "dependent" and "independent" variables, respectfully, hence, the choice of indices. Once the binding constraints are identified, it suffices to minimize $f[x_D(x_I),0]$ over x_I using any appropriate <u>unconstrained</u> method, to determine x_I^*, and hence x^*. The indicated algorithms actually invoke the linear independence assumption for <u>all</u> feasible boundary points, and hence at any given iteration can either reduce $f(x)$ without encountering constraints, or will be in a situation completely analogous to the one described at the outset, and can proceed to minimize f over the currently independent variables in the space of currently binding constraints.

Returning to the determination of S and $P^\#$ for this type of algorithm, we observe that $S \equiv (S_D^T, S_I^T)^T$ must satisfy $PS = (P_D, P_I)(S_D^T, S_I^T)^T = 0$, so $S_D = -P_D^{-1} P_I S_I$ and hence

$$S = \begin{pmatrix} -P_D^{-1}P_I \\ I \end{pmatrix} S_I .$$

Similarly, since $PP^\#P = P$, defining $P^\# \equiv ((P_1)^T, (P_2)^T)^T$ gives the result

$$P^\# = \begin{pmatrix} P_D^{-1}(I-P_IP_2) \\ P_2 \end{pmatrix} = \begin{bmatrix} P_D^{-1} \\ 0 \end{bmatrix} + \begin{bmatrix} -P_D^{-1}P_I \\ I \end{bmatrix} P_2 .$$

In terms of the quantities defined, the block components A_{ij} of $\bar{M}^1$ are therefore given by (5.10), where

$$S = T\,S_I\ ,\quad P^{\#} = \begin{bmatrix} P_D^{-1} \\ 0 \end{bmatrix} + T\,P_2\ ,$$

and $T = \begin{bmatrix} -P_D^{-1}P_I \\ I \end{bmatrix}$, an n by $[n-(r+p)]$ matrix with rank $[n-(r+p)]$, S_I is any $[n-(r+p)]$ square nonsingular matrix, and P_2 is any $[n-(r+p)]$ by $(r+p)$ matrix.

For the projected gradient type algorithms, the gradients of binding constraints are again assumed linearly independent at feasible boundary points, with the data being organized mainly to accommodate a projection matrix of the form, $P_R = I - P^T(PP^T)^{-1}P$, used to project a given direction vector into the linear subspace associated with the currently binding constraints. Here, the rows of P would represent the gradients of the constraints currently deemed to remain binding in the next iteration, and PP^T is an $(r+p)$ by $(r+p)$ matrix, nonsingular under the linear independence assumption. We find that $P^{\#} = P^T(PP^T)^{-1}$ satisfies the requirements for a pseudo-inverse of P, and it follows easily that a suitable choice for S is any matrix S_R formed by selecting any $[n-(r+p)]$ linearly independent columns of P_R. The block components of $\bar{M}^1$ for gradient projection type calculations are therefore the same as (5.10), with $S = S_R$ and $P^{\#} = P^T(PP^T)^{-1}$.

Therefore, returning to the calculation of the derivatives of the Kuhn-Tucker triple we evaluate (5.6) for the representation (5.7) of $\bar{M}^1$ and the expression (5.4) of $\bar{N}$ to obtain

$$\nabla_\varepsilon \bar{y}(\varepsilon) = \begin{bmatrix} \nabla_\varepsilon x(\varepsilon) \\ \hline \nabla_\varepsilon \bar{u}(\varepsilon) \\ \nabla_\varepsilon w(\varepsilon) \end{bmatrix} = \bar{M}(\varepsilon)^{-1}\, \bar{N}(\varepsilon) = \left[\begin{array}{c} -A_{11} \dfrac{\partial(\nabla_x L^T)}{\partial \varepsilon} + A_{12} \begin{bmatrix} \dfrac{\partial \bar{g}}{\partial \varepsilon} \\ -\dfrac{\partial h}{\partial \varepsilon} \end{bmatrix} \\ \hline -A_{21} \dfrac{\partial(\nabla_x L^T)}{\partial \varepsilon} + A_{22} \begin{bmatrix} \dfrac{\partial \bar{g}}{\partial \varepsilon} \\ -\dfrac{\partial h}{\partial \varepsilon} \end{bmatrix} \end{array}\right], \quad (5.11)$$

the A_{ij} being given by (5.8), (5.9) or (5.10), depending on the respective conditions that apply and depending on how the data are organized.

The Hessian $\nabla_\varepsilon^2 f^*(\varepsilon)$ (4.5) of the optimal value function may also be readily calculated once (5.11) has been evaluated. To do this efficiently, first note that we may rewrite (4.5) as

$$\nabla_\varepsilon f^*(\varepsilon) = \partial^2 L/\partial \varepsilon^2 + [(\partial(\nabla_x L^T)/\partial \varepsilon)^T, -(\partial g/\partial \varepsilon)^T, (\partial h/\partial \varepsilon)^T]\, \nabla_\varepsilon y(\varepsilon) . \quad (5.12)$$

Denoting by $\overline{\nabla_x^2 f^*}$ the "reduced Hessian" that results from eliminating terms associated with nonbinding constraints, and using the previous notation, we obtain the concise expression,

$$\overline{\nabla_\varepsilon^2 f^*}(\varepsilon) = \partial^2 L/\partial \varepsilon^2 - \bar{N}^T \nabla_\varepsilon \bar{y}(\varepsilon) , \quad (5.13)$$

or equivalently,

$$\overline{\nabla_\varepsilon^2 f^*}(\varepsilon) = \partial^2 L/\partial \varepsilon^2 - \bar{N}^T \bar{M}^{-1} \bar{N} . \quad (5.14)$$

The Hessian can now be calculated from the given problem data and (5.13) or (5.14), using (5.11), evaluating the A_{ij} as given in (5.8), (5.9) or (5.10), depending on which conditions apply.

For Problem R(ε), (5.11) simplifies considerably, as shown by Armacost and Fiacco [5]. It is easy to verify that for this problem,

$\partial(\nabla_x L^T)/\partial\varepsilon = 0$, $\partial\bar{g}/\partial\varepsilon = [-I,0]$ and $\partial h/\partial\varepsilon = [0,-I]$. Therefore, (5.11) becomes

$$\nabla_\varepsilon \bar{y}(\varepsilon) = \begin{bmatrix} \nabla_\varepsilon x(\varepsilon) \\ \hline \nabla_\varepsilon \bar{u}(\varepsilon) \\ \nabla_\varepsilon w(\varepsilon) \end{bmatrix} = \begin{bmatrix} A_{12} \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \\ \hline A_{22} \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \end{bmatrix} \tag{5.15}$$

The general formulas for $\nabla^2 f^*(\varepsilon)$ also simplify for Problem $R(\varepsilon)$. Observe that for this problem, $\nabla^2 L/\partial\varepsilon^2 = 0$ and $\bar{N} = \left[0 \ , \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}\right]^T$. Hence, using the form (5.13) we obtain

$$\overline{\nabla^2 f^*}(\varepsilon) = -\left[0 \ , \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}\right] \begin{bmatrix} \nabla_\varepsilon x \\ \nabla_\varepsilon \bar{u} \\ \nabla_\varepsilon w \end{bmatrix} = \begin{bmatrix} \nabla_\varepsilon \bar{u} \\ -\nabla_\varepsilon w \end{bmatrix},$$

which essentially agrees with the result obtained in Corollary 4.1, and using the form (5.14) we obtain the interesting result,

$$\overline{\nabla^2 f^*}(\varepsilon) = -\begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} A_{22} \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}. \tag{5.16}$$

Aside from the considerable computational simplification compared to the general problem, these results provide additional insights into the structure of the solution, since we have explicit relationships for the various parameter-derivatives in forms of quantities associated with the original problem functions. For example, noting the result (5.16) and the various possibilities for A_{22} given in (5.8), (5.9) or (5.10), we can see directly that $f^*(\varepsilon)$ (associated with $R(\varepsilon)$) is convex in a neighborhood of $\varepsilon = 0$ if the Lagrangian $L(y,\varepsilon)$ of $R(\varepsilon)$ is convex in x. This well known fact and several related and less well known

inferences associated with Problem R(ε) can be shown explicitly using the given formulas. For additional details and results, see the paper by Armacost and Fiacco [5].

6. Discussion of Results and Extensions

The nontrivial computational considerations associated with checking whether all the conditions are satisfied, as required by the assumptions of Theorem 3.1 and further compounded by the refinements associated with the appropriate calculation of the A_{ij}, are typical of analogous verification problems confronted by most numerical procedures. Additional difficulties are, of course, associated with the (typical) requirement to essentially know the solution before the conditions required to solve the problem can be verified or the solution analyzed. Such concerns are outside the scope of the presentation, that is primarily concerned with the existence and characterization of relationships that hold at a solution. However, a few relevant comments can be offered.

As stated briefly at the outset, a method for estimating solution sensitivity information by using penalty function methods was established by Fiacco [15], implemented on the computer by Armacost and Mylander [8] and extended and applied by Armacost and Fiacco [3] - [6]. This approach is based on the fact that the local solution matrix of first partial derivatives $\nabla_\varepsilon x(\varepsilon)$, the optimal value $f^*(\varepsilon)$ and the gradient $\nabla_\varepsilon f^*(\varepsilon)$ and Hessian $\nabla^2_\varepsilon f^*(\varepsilon)$ of the optimal value function, are component by component limits of the parameter-derivatives of the penalty function minimizing point, optimal penalty function (parameter) gradient and Hessian, respectively, under the given conditions. In effect, a class of algorithms was shown to generate a trajectory that both terminates at a solution and rather faithfully reflects the perturbation behavior (subject to parameter changes) of the solution, as the solution is approached. Furthermore, the calculations required to determine the sensitivity information turn out to be of the same form as the calculations required by the algorithm to generate a solution trajectory. Thus, for such algorithms applied to problems

satisfying the appropriate conditions, increasingly accurate estimates of the sensitivity information is available with little extra effort, as the solution is approached, i.e., the solution need not be known in advance of easily determining certain aspects of its behavior. (Another example of this sort of result is the calculation of error bounds in solving systems of equations. For an application to nonlinear programming, see the paper by Robinson [22].)

A second theoretical advantage of the penalty function approach to estimating solution sensitivity involves the calculation of the A_{ij} defining $\bar{M}^{-1}$ (5.7) and is worth noting. Under the assumed conditions, the Hessian of the penalty function is positive definite near a solution. (See Fiacco and McCormick [16, Theorem 12] and Fiacco [15]). The stationarity condition of the penalty function at the minimizing point essentially approximates (with appropriate interpretation) the information given in (3.1), and the result is a single formula (obtained by Fiacco [15]) for the approximation of $\nabla_\varepsilon y(\varepsilon)$. Thus, there are no alternative calculations such as (5.8), (5.9), or (5.10) that depend on the status of the solution. (Armacost and Fiacco [4] provide a detailed treatment of the penalty function estimates.)

The latter advantage has been shown to extend to augmented Lagrangian functions by Armacost and Fiacco [7], Armacost [2], and Buys and Gonin [13]. Indeed, it is clear that unique formalas for $\nabla_\varepsilon y(\varepsilon)$ will obtain for that family of generalized Lagrangians and exact penalty functions that are structured such that their Hessians are positive definite at a Kuhn-Tucker triple under the conditions of Theorem 3.1. A large class of such Lagrangians was developed by Arrow, Gould and Howe [9]. Essentially, if the extended Lagrangian is denoted by ϕ then, since the role of ϕ is precisely analogous to the role of the usual Lagrangian L , and since $\nabla_x \phi = 0$ and $\nabla_x^2 \phi$ is positive definite at a Kuhn-Tucker triple, it follows that ϕ can replace L in the results given. In particular, it follows that the A_{ij} that determines $\bar{M}^{-1}$ are uniquely given by (5.8), with ϕ replacing L in those formulas.

It is also clear that, if the Lagrange multipliers (of such an extended Lagrangian) are sufficiently smooth functions of the problem parameters that converge to (locally) optimal multipliers, then the associated minimizing point of the Lagrangian function, along with the parameter-derivatives of the minimizing point, will converge respectively to $x(\varepsilon)$ and $\nabla_\varepsilon x(\varepsilon)$, and the optimal value, gradient and Hessian of the Lagrangian will converge to $f^*(\varepsilon)$, $\nabla_\varepsilon f^*(\varepsilon)$ and $\nabla^2_\varepsilon f^*(\varepsilon)$, respectively. Thus, these functions also give rise to techniques for estimating sensitivity information prior to the determination of a solution, analogous to those obtained for penalty functions. An extension of the class of algorithms that can be so utilized should continue to be an interesting subject of future research.

In the other direction, that of using solution sensitivity information to characterize algorithmic behavior, interesting examples are the proof by Meyer [21] of convergence of a family of algorithms and the determination by Robinson [22] of the convergence and rate of convergence of a large class of algorithms.

Finally, though we have concentrated on sensitivity analysis and developed neighborhood results, some of these results may be expected to extend to parametric nonlinear programming, where the parameters are permitted to range in a prescribed set. A characterization and sensitivity and stability analysis of parameter-dependent solutions will undoubtedly be a subject of sustained future investigation. It seems apparent that results "in the large" will depend critically on neighborhood results such as those presented here.

An immediate application of the sensitivity analysis results obtained here is a calculation of first order estimates of a Kuhn-Tucker triple of a problem with parameter changes, and first and second order estimates of the optimal value function, using Taylor's series expansions. If $x(0)$ is a solution of Problem P(0) satisfying the conditions of Theorem 3.1, then a first order estimate of the optimal value function,

$f^*(\varepsilon) \equiv f[x(\varepsilon),\varepsilon]$, for ε in a neighborhood of 0 , is given by

$$f^*(\varepsilon) = f^*(0) + \nabla_\varepsilon f^*(0)\varepsilon \ , \tag{6.1}$$

and a second order estimate is given by

$$f^*(\varepsilon) = f^*(0) + \nabla_\varepsilon f^*(0)\varepsilon + \frac{1}{2}\varepsilon^T\nabla_\varepsilon^2 f^*(0)\varepsilon \ , \tag{6.2}$$

where $f^*(0) = f(x^*,0)$ and $\nabla_\varepsilon f^*(\varepsilon)$ and $\nabla_\varepsilon^2 f^*(\varepsilon)$ are defined by (4.4) and (4.5), respectively. A first order estimate of the Kuhn-Tucker triple $y(\varepsilon)$ is given by

$$\begin{aligned} y(\varepsilon) &= y(0) + \nabla_\varepsilon y(0)\varepsilon \\ &= \begin{bmatrix} x^* \\ u^* \\ w^* \end{bmatrix} + M^{-1}(0)N(0)\varepsilon \ , \end{aligned} \tag{6.3}$$

where we have used Conclusion (1) of Theorem 3.1 and Equation (3.3).

An ever important general application of sensitivity analysis is the determination of those parameters to which a solution is the most sensitive. In the context of mathematical programming, if the optimal value or one or more components of a solution vector or any of the constraints can change erratically for small changes in a parameter, there is little comfort in having a particular solution at hand for the given data, if the data is (as usual) subject to errors or alterations that can exceed these "small changes." A sensitivity analysis can thus lead to the more likely sources of instability in the model and to a further study of data inaccuracy (e.g., suggesting more observations to reduce the variance of sample estimates, as in a chance constrained formulation of a problem studied by Armacost and Fiacco [3]). It can also suggest reformulating the model to eliminate various instabilities (e.g., by refraining from expressing an equality constraint as two inequalities, the consequences of which are easily seen to make singular the Jacobian M of the Kuhn-Tucker system (3.1), the computational implications being dramatically conveyed by Robinson [23]) or introducing "regularizations,"

i.e., estimates or perturbations that introduce stabilities (e.g., replacing nondifferentiable functions by differentiable approximations or perturbing a function so that the Hessian of the perturbed function is nonsingular in an appropriate domain, as in the various definitions of augmented Lagragians).

The sensitivity information for the optimal value function and the Kuhn-Tucker triple can also be used to drive various "cyclic" procedures for solving problems involving optimization, e.g., in solving $\min_x \max_y F(x,y)$ by cycling between descent moves in x-space and ascent moves in y-space, where the parameter ε of $P(\varepsilon)$ would essentially momentarily correspond to that subset of variables that are considered to be "independent" for a given iteration. An excellent discussion of this sort of method may be found in a paper by Hogan [17] and a recent application using penalty function approximations mentioned earlier and validated in [15] was given by de Silva [14]. The latter involves the solution of an implicitly defined optimization model of U.S. crude oil production.

For Problem $R(\varepsilon)$, where the parameters are the right-hand side of the constraints, the Kuhn-Tucker triple derivatives (5.15) and the Hessian of the optimal value function (5.16) are relatively easy calculations and should have powerful application in solving large-scale problems by introducing Newton-type techniques in the various established decomposition procedures. Problems of this type are also intimately involved in much of duality theory and sensitivity information may have useful application in defining and accelerating algorithms for solving $R(0)$ by various dual methods. Sensitivity results for Problem $R(\varepsilon)$ are treated in considerable detail by Armacost and Fiacco [5]. Potential applications are abundant.

We have presented a number of basic results for a locally rather ideally behaved class of nonlinear programming problems. Results involving the general behavior of the optimal value function and a given solution or solution set, under less stringent assumptions, have been known for some time, and numerous significant refinements, extensions, and generalizations have been obtained only recently. The subject of sensitivity and stability

analysis in nonlinear programming is finally receiving the attention it deserves. The reader interested in pursuing the subject further may make an excellent start by studying the articles [17], [18], and [19] by Hogan and consulting the numerous references therein.

REFERENCES

[1] ARMACOST, ROBERT L. (1976). Computational Experience With Optimal Value Function and Lagrange Multiplier Sensitivity. Technical Paper Serial T-335. Institute for Management Science and Engineering, The George Washington University.

[2] ARMACOST, ROBERT L. (1976). Sensitivity Analysis in Parametric Nonlinear Programming. Doctoral Dissertation, School of Engineering and Applied Science, The George Washington University.

[3] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1974). Computational experience in sensitivity analysis for nonlinear programming. Mathematical Programming 6 301-326.

[4] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1975). Second-Order Parametric Sensitivity Analysis in NLP and Estimates by Penalty Function Methods. Technical Paper Serial T-326. Institute for Management Science and Engineering, The George Washington University.

[5] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1976). NLP sensitivity for R.H.S. perturbations: A brief survey and recent second-order extensions. Technical Paper Serial T-334. Institute for Management Science and Engineering, The George Washington University.

REFERENCES

[1] ARMACOST, ROBERT L. (1976). Computational Experience With Optimal Value Function and Lagrange Multiplier Sensitivity. Technical Paper Serial T-335. Institute for Management Science and Engineering, The George Washington University.

[2] ARMACOST, ROBERT L. (1976). Sensitivity Analysis in Parametric Nonlinear Programming. Doctoral Dissertation, School of Engineering and Applied Science, The George Washington University.

[3] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1974). Computational experience in sensitivity analysis for nonlinear programming. Mathematical Programming 6 301-326.

[4] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1975). Second-Order Parametric Sensitivity Analysis in NLP and Estimates by Penalty Function Methods. Technical Paper Serial T-326. Institute for Management Science and Engineering, The George Washington University.

[5] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1976). NLP sensitivity for R.H.S. perturbations: A brief survey and recent second-order extensions. Technical Paper Serial T-334. Institute for Management Science and Engineering, The George Washington University.

[6] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1976). Sensitivity Analysis for Parametric Nonlinear Programming Using Penalty Methods. Technical Paper Serial T-340. Institute for Management Science and Engineering, The George Washington University, Washington, D. C.

[7] ARMACOST, ROBERT L. and ANTHONY V. FIACCO (1977). Exact Sensitivity Analysis Using Augmented Lagrangians. Technical Paper Serial T-349. Institute for Management Science and Engineering, The George Washington University.

[8] ARMACOST, ROBERT L. and W. CHARLES MYLANDER (1973). A guide to a SUMT-Version 4 computer subroutine for implementing sensitivity analysis in nonlinear programming. Technical Paper Serial T-287. Program in Logistics, The George Washington University.

[9] ARROW, K. J., F. J. GOULD, and S. M. HOWE (1973). A General Siddle Point Result for Constrained Optimization. Mathematical Programming 5 225-234.

[10] BERTSEKAS, D. P. (1973). On penalty and multiplier methods for constrained minimization. EES Department Working Paper (August). Stanford University.

[11] BIGELOW, JAMES H. and NORMAN Z. SHAPIRO (1974). Implicit function theorems for mathematical programming and for systems of inequalities. Mathematical Programming 6 141-156.

[12] BUYS, J. D. (1972). Dual algorithms for constrained optimization problems. Doctoral Dissertation, University of Leiden, Rotterdam, Netherlands.

[13] BUYS, J. D. and R. GONIN (1977). The Use of Augmented Lagrangian Functions for Sensitivity Analysis in Nonlinear Programming. Mathematical Programming 12 281-284.

[14] DE SILVA, A. H. (1978). Sensitivity Formulas for Nonlinear Factorable Programming and Their Application to the Solution of an Implicitly Defined Optimization Model of U.S. Crude Oil Production. Doctoral Dissertation, School of Engineering and Applied Science, The George Washington University.

[15] FIACCO, ANTHONY V. (1973). Sensitivity analysis for nonlinear programming using penalty methods. Mathematical Programming 10 287-311.

[16] FIACCO, ANTHONY V. and GARTH P. McCORMICK (1968). Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York.

[17] HOGAN, WILLIAM (1973). Directional Derivatives for Extremal-Value Functions With Applications to the Completely Convex Case. Operations Res. 21 188-209.

[18] HOGAN, WILLIAM (1973). The Continuity of the Perturbation Function of a Convex Program. Operations Res. 21 351-352.

[19] HOGAN, WILLIAM (1973). Point-to-Set-Maps in Mathematical Programming. SIAM Rev. 15 591-603.

[20] McCORMICK, GARTH P. (1977). Future Directions. Mathematical Programming. NATO-ASI Lecture Notes, Engineering Plasticity by Mathematical Programming, University of Waterloo, Waterloo, Ontario, Canada.

[21] MEYER, R. (1970). The Validity of a Family of Optimization Methods. SIAM J. CONTROL 3 (No. 1).

[22] ROBINSON, STEPHEN M. (1974). Perturbed Kuhn-Tucker points and rate of convergence for a class of nonlinear-programming problems. Mathematical Programming 7 1-16.

[23] ROBINSON, S. M. (1977). A Characterization of Stability in Linear Programming. Operations Res. 25 435-447.